

Table of Contents

Introduction	1
Key Differences from the 2008 Norms	1
Using 2008 Data	2
Kindergarten Norms	3
Science Norms for Grade 2	4
Norms for Grades 11 and 12	4
Median Scores	4
Using Norms for Student Placement	5
Methodology	5
Where to Access Norms Resources	5

Introduction

NWEA conducts norming studies every three years, and the most recent NWEA norms were released in July 2011. It is an important part of our commitment to partnership: Providing the best information we can about student achievement and growth to better inform educational decision-making.

The methodological changes used to create the 2011 norms have resulted in a set of norms that are more representative of the US school age population than previous NWEA norms. As a result of the improved methodology, partners can expect some differences from the 2008 norms, as explained below.

Key Differences from the 2008 Norms

Will I see changes on my reports?

All partners will see some change. The status and growth norm tables have been updated to reflect the 2011 norms and new norms have been applied to both current and past results. Reporting all results in relation to a single set of norms allows partners to accurately evaluate trends in performance and improvement over time.

What changes will I see in the *status norms*?

Changes in the status norms are minor. As in the new status norms, status norms for 2008 also refer to the US school-age population. In general, the percentile rankings associated with student scores changed very little between the 2008 and 2011 norms. Partners may observe somewhat larger changes in grades 1 and 2, specifically for the low and high performing students. The differences reflect our changing partner base.

What changes will I see in the growth norms?

Changes in the growth norms are minor in the middle of the achievement continuum, and larger at the ends of the achievement continuum, particularly at the very extremes. These changes are the result of new methods to improve the representativeness of the norming sample and the introduction of the new estimation procedures. Students who are very high or very low achievers may see significant changes in their growth estimates. For low achievers, they will typically see a somewhat lower growth estimate than they have in the past. For high achievers, the growth estimate will often increase.

The largest changes occur at the most extreme initial scores. This means that relatively small numbers of students will see very large changes. Therefore, classrooms and schools with large numbers of very low performing or very high performing students may see larger changes in their growth estimates.

Using 2008 Data

Why did you apply the new process to all the old information? Can we access the old data as an option?

When comparing achievement data across time, it is essential that the data are being compared to a single norm. Since the 2011 norms better represent the U.S. school age population and give more robust estimates of growth across all ranges, we believe the 2011 norms provide a more interpretable comparison group, not only for 2011 student test results, but also for prior results.

There are circumstances where schools will need to produce reports that show comparisons relative to the 2008 norms. In these cases, we will need to work with the partner to understand the problem and help find a solution that will address their needs.

According to the new norms, my school showed less growth than was reported by the prior norms. Why?

Your school is no different than it was before, but the norm group did change, and that subsequently changed your growth results relative to this new group. It also means that comparing your 2011 results to your 2008 results is somewhat like comparing apples to crabapples. On its face it may seem like these results can be compared, but they really should not be.

What's more important is where these new growth norms place you relative to your prior results. When we updated the status and growth norms, we updated them for each and every testing term. That means your 2008, 2009, and 2010 results can be reproduced to reflect the new norms. This allows you to make apples to apples comparisons that allow you to fairly evaluate trends in achievement and growth in your system over time.

Growth Targets

Do all of the students in a particular grade level have the same growth target? Can growth be flat in all grade levels/subjects?

No. The only grades in which the norming sample shows flat average growth are in grades 5, 6, and 8 in mathematics. The new norming study changed the sampling method and used all of the available data to estimate the growth norms for students on the ends of the scale. The new estimates suggest that school age children will, when averaged over a very large nationally representative group, generally show nearly the same growth at all points of the RIT scale, all else being equal.

It is true that average growth projections are flatter in the new norming study than in prior norming studies, and this is also a product of the change in sampling methodology. The overly strong influence of students from both extremes of the starting RIT range in the 2008 growth norms appear to be appropriately moderate in the 2011 norms.

Should we use the growth norms to create growth targets?

The growth norms were never intended as growth targets, they simply represent our best estimate of the average growth of students at the various points of the RIT scale. In other words, they represent the growth of the proverbial "middle child" in the norming population. Target growth should be the growth that the student and family aspire to achieve. The growth norms can provide some parameters for discussion by helping the teacher, child, and family understand the range of growth that's typical.

Standard Errors of Measurement (SEM)

How can partners accurately assess student growth compared to typical growth when the typical growth is less than the Standard Error of Measure (SEM)?

It is also important to keep in mind that there are several kinds of standard errors in play: the standard error of the predicted growth (which is relatively miniscule due to the large samples employed and can be ignored for practical purposes), the SEMs of each of the observed scores (e.g., fall and spring), and the SEM associated with the observed growth (e.g., spring – fall). It is the SEM associated with the observed growth that needs to be considered when evaluating growth. This value is typically around 4.3 for two test occasions that each functioned well for a student. It is not wrong or irrelevant to evaluate observed growth that is less than typical growth, even when typical growth is less than 4 RITs.

Growth Percentage Changes

How come my students' growth percentage has dropped by 10+% in DRS? If you've changed your methodology, why does it need to be applied to previous terms? Why is proficiency exactly the same, and growth so different?

It is important to understand that the change from 2008 to 2011 norms has no bearing on your students' data. The growth your students made last school year has not changed. What has changed is the *norming comparison*. And because the methods for calculating the 2008 and 2011 norms are quite different, it is inappropriate to compare results using the 2008 norms to those using 2011 norms.

The reporting system has recalculated each student's past history to reflect the 2011 norms. So teachers and schools can compare their current growth and performance to their prior growth and performance using the 2011 norms. This assures all of your data has been updated to reflect a norm group comparison that is more representative of the U.S. school age population and a more robust estimate of growth across all parts of our scale.

So if you were to look at the difference between the growth of this year's third graders relative to last year's third graders, our current reports show you those results in relation to the 2011 norms, and these provide better data for such a comparison.

Kindergarten Norms

What status norms are available for kindergarteners?

Fall, winter, and spring status norms are available now and will appear on your reports.

NWEA employed two different methods for calculating status norms for kindergarten based upon the tested season. The 2008 norm methodology was used for the fall and winter terms because these are baseline test seasons for students. We were able to employ the 2011 methodology to calculate the spring status norm because more testing information such as student test history was available. One word of note, both of these methods however, differ from other grades in that we do not adjust for the unique kindergarten instruction time differences, i.e. ½ day vs. full day.

What growth norms are available for kindergarteners?

Fall-to-winter and fall-to-spring growth norms are no longer available, but spring-to-spring growth norms are now available. Many schools offer kindergarteners only a half-day of instruction and, unfortunately, the amount of daily instructional time received by kindergarteners is not reported to us by our partner schools. Because this information is incomplete, the resulting average would not be accurate for either group.

Spring-to-spring growth norms are available – from spring of kindergarten to spring of 1st grade – as the majority of growth in this period occurs during the 1st grade and is not significantly affected by the kindergarten schedule.

Why should kindergarteners be tested for fall if there is nothing to offer them besides RIT score and goals?

Districts should not be using our test only to get a percentile score that compares the student to peers. The percentile score, while important, is not the primary reason to test a student. The RIT score provides a baseline from which future growth can be tracked. Our MPG tests offer checklists that allow teachers to see the particular strengths and weaknesses of all students in their classrooms, and the more advanced MPG tests provide useful information through the goal scores and Primary Grades Instructional Data.

Science Norms for Grade 2

Where are the status and growth norms for 2nd grade General Science and Concepts & Processes?

Status norms for 2nd grade General Science and Science Concepts & Processes are not included. Grade 2 students are tested in science much less frequently and as a much lower proportion of all grade 2 students than they are for reading, mathematics, and even language usage. Given these low testing/participation levels, the grade 2 sample would not be representative of all of the grade 2 students that NWEA tests.

One of the major goals of the 2011 norming study was to sample in a manner that better assured norms that reflected the U.S. school age population. When that goal could not be met because of insufficient levels of testing, we chose to not report a norm rather than report a norm that violated this principle.

Growth norms for the spring of 2nd grade to the spring of 3rd grade are available as the majority of this growth occurs during the 3rd grade academic year.

Norms for Grades 11 and 12

Why are there no norms or percentiles for 11th and 12th grades?

Status norms for 11th and 12th grades and growth norms for spring of 10th grade to spring of 11th grade are not included because we concluded that insufficient numbers of students tested to permit creation of a robust, representative sample. The number of students tested with MAP drops off substantially after grade 10, suggesting that testing practices for grades 11 and 12 are varied within and across schools and districts.

To put it simply, when we had reason to believe that a norm would not be representative of the U.S. school-age population and their performance or growth distribution, we chose not to report a norm, rather than report a norm that might be inaccurate or misleading.

Median Scores

Why is the median RIT score missing from the new norms study?

By definition, the median (or midpoint) of a normal distribution is the same as the mean. Since that is the case, the mean score reported in our norms would also represent the median score. Of course, the median in any classroom might not be similar to the mean, so we will continue to report median scores in teacher reports and other reports.

Calculating Instructional Time

For a newly enrolled student, testing for the first time with MAP, how will “Instructional Time” be calculated for that child?

If the student has no prior history of testing with MAP, then something simple could be used to estimate the instructional weeks prior to fall testing (like the average of the other students’ instructional weeks).

The reporting system makes assumptions about instructional time based on the data gathered and uses default values when presenting status and growth norms. Fall is approximately 4 weeks from the beginning of the academic year, winter is approximately 17 weeks, and spring approximately 30 weeks, although these default values vary slightly depending on the grade and subject.

Using Norms for Student Placement

Since the change in the norms, several students who did not qualify for gifted placement last spring under the old norms now have percentile rankings associated with that same score that would qualify them. What do we do?

Our recommendation is always to give the benefit of the doubt to the student in these circumstances, particularly because the new status norms were created using a better sampling and estimation procedure. We would recommend treating all students who achieved a qualifying score, whether that score was gained under the old norms or the new norms, as crossing the qualifying threshold. And of course we'd recommend that the district use multiple measures in addition to NWEA MAP scores to determine gifted program eligibility.

Methodology

For the 2011 norms, students were randomly sampled from the pool of test records of the **5.1 million students** who took MAP tests in **spring and fall terms of 2009 and 2010, as well as the winter term of 2010.**

Students represented in the pool came from **over 13,000 schools in more than 2,700 school districts in 50 states.** For each student record selected, all of the student's test events from the target terms (that met basic quality and validity criteria) were also retrieved.

Using data from the National Center for Education Statistics (NCES) database, a regression procedure was used to order schools in terms of the economic challenges and opportunities their students face. Information contained in the resulting **School Challenge Index (SCI)** was then used to derive post-stratification weights for NWEA schools. *These weights are essential to building norms that are representative of and therefore generalize to all school-age students served by public schools in any particular state.*

This allows us to present our current and potential partners with **comparisons to other schools in their state** that have similar levels of challenge. In addition, the sample for the 2011 norms is stratified to the national school-age population so that **state-to-national comparisons are relevant as well.**

The 2005 and 2008 NWEA norms were based on a different methodology. The 2008 norms were developed using all test records from a five-year period that met all basic quality and validity criteria.

Where to Access Norms Resources

On the Reports landing page of the NWEA website you will find links to a variety of norms-related documents.

The 2011 RIT Scale Norms Study will be available in September and linked from the Reports landing page. The study will be available in digital format only, and partners may print as many pages as they like, as often as needed.

Have a question? Please contact your NWEA Partner Relations Representative at 503-624-1951 or NWEA Technical Support at 877-469-3287.